Motion Imagery Standards Board **Recommended Practice** **Real-Time Protocol for Full Motion Video**	**MISB RP 0804** **25 March 2008**

1 Scope

This Recommended Practice (RP) documents the standards profile for packaging and delivering Full Motion Video (FMV as defined in MISP 4.5) data over the Real-Time Protocol (RTP). This RP provides direction on the packetization and streaming of video and metadata using RTP to support diverse IP based networks.

The scope of this RP is limited to delivery of Full Motion Video products and is not intended to replace any other approved standards for other uses; rather it is intended to complement those standards.

2 References

2.1 Normative References

[1] ISMA 2.0, *Internet Streaming Media Alliance Implementation Specification,* April 2005

[2] ISO/IEC 14496-10:2005, *Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding* ITU-T Recommendation H.264 *Advanced video coding for generic audiovisual services*, November 2007

[3] MISB EG 0802, *H.264/AVC Coding and Carriage*, July 2008

[4] MISB EG 0104.5, *Predator UAV Basic Universal Metadata Set*, May 2008

[5] MISB EG 0601.1, *UAS Datalink Local Metadata Set*, December 2007

[6] MISB RP 0604, *Time Stamping Compressed Motion Imagery*, June 2007

[7] SMPTE 336M-2001, *Data Encoding Protocol Using Key-Length-Value*

[8] MISB EG 0701 *Common Metadata System: Structure,* August 2007

[9] RFC 2326, *Real Time Streaming Protocol (RTSP)*, April 1998

[10] RFC 2327, *SDP: Session Description Protocol*, April 1998

[11] RFC 3550, *RTP: A Transport Protocol for Real-Time Applications*, July 2003

[12] RFC 3551, *RTP Profile for Audio and Video Conferences with Minimal Control*, July 2003

[13] RFC 3984, *RTP Payload Format for H.264 Video*, February 2005

2.2 Informative References

SMPTE 335M-2001, *Metadata Dictionary Structure*

SMPTE RP210.8-2004, *Metadata Dictionary*

SMPTE RP210.7-2003, *Metadata Dictionary*

SMPTE RP210.3-2001, *Metadata Dictionary* (DRAFT)

MISP 4.5, *Motion Imagery Standards Profile*, May 2008

MISB RP 0101, *Use of MPEG-2 Systems Streams in Digital Motion Imagery Systems*

MISB RP 0103.1, *Timing Reconciliation Metadata Set for Digital Motion Imagery*,

11 October 2001

MISB RP 0107, *Bit and Byte Order for Metadata in Motion Imagery Files and Streams*, 11 October 2001

MISB TRM-07A, *Low Bandwidth Motion Imagery – Technologies*, November 2007

3 Acronyms

FMV	Full Motion Video
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
ISMA	Internet Streaming Media Alliance
MPEG2-PES	MPEG2 compressed video packetized elementary stream
MPEG2-TS	MPEG2 Transport Stream
MP4/AVC	MPEG4 Part 10 AVC/H.264 (a.k.a H.264/AVC)
NALU	Network Abstraction Layer Unit
RTP	Real Time Protocol
RTP/AVP	IETF's RTP using Audio/Video profile carried over UDP
RTCP	Real Time Control Protocol
RTSP	Real Time Streaming Protocol
SDP	Session Description Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
URL	Uniform Resource Locator

4 Introduction

This RP defines the standards profile used to provide FMV to users via RTP over IP (Internet Protocol) based networks. The scope of what this RP is attempting to provide is very broad, and given the wide variety of infrastructure, client device, user requirements, and other considerations, it does not attempt to specifically address all possible permutations. Rather, this RP focuses on the standards that provide broad flexibility, and reference implementation guidance for RTP to determine how to best apply it for specific needs.

4.1 RTP FMV Features

RTP has been designed to accommodate the nuances of internet-centric multimedia streaming. It offers the following capabilities:

- Delivery of digital motion video over various network and link types that may exhibit packet loss, packet re-ordering, latency, and jitter.
- A standardized method for requesting an RTP stream from a digital motion video provider.
- A standardized method for stream control to allow trick play.
- Authentication and encryption of data to provide integrity, confidentiality and non-repudiation.
- Provisions for lowering the overhead associated with packetizing and streaming data.

4.2 Relationship with established Internet Standards

The Internet provides a good example of the challenges faced when delivering data over disparate best-effort networks of varying qualities. ISMA [1] defines a set of standards for storage and streaming of media over the Internet; this RP aligns itself with that family of standards.

4.3 Relationship with established MISB Standards

The following figure illustrates the relationships between the current MISP standards and those in this document.

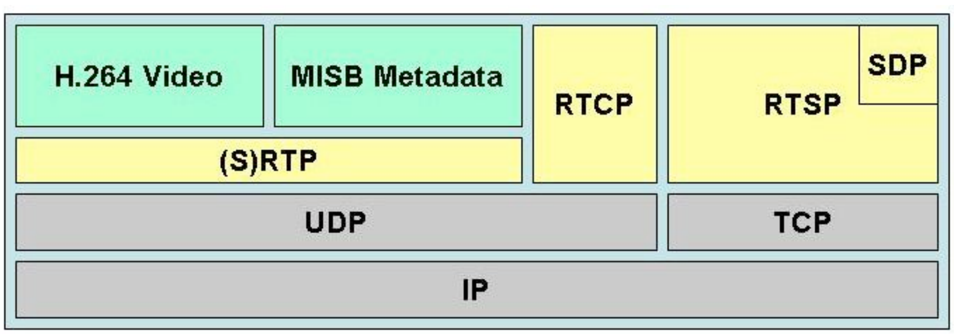

Figure 1. RTP Relationship with Current MISP Standards

Low-bandwidth digital motion imagery is intended to align with established MISP standards where appropriate; however, more optimal protocols are emerging for Internet use. Evaluation and adoption of new standards for more robust delivery is encouraged as part of the regular activity of the MISB.

5 Architecture

The high level architecture consists of a data provider or source of FMV, and a data receiver of the FMV. Bidirectional data flow is required to enable full functionality, such as channel adaptation and trick play. When such control of the stream is warranted protocols such as RTSP (Real Time Streaming Protocol), SDP (session Description Protocol), and RTCP (Real Time Control Protocol) are necessary. However, simpler configurations for direct video and/or metadata playback only can be done with RTP alone.

The term “client” is used to refer to the endpoint receiving data from some data producer; this can be an end user (e.g., Warfighter) or another system.

6 Data Formats

To provide broad applicability for devices that may only need one media component of the available data a client SHALL support both video and metadata delivery and MAY support audio. This RP will address video and metadata only.

6.1 Video

The required codec in this document is MPEG4 Part 10 AVC/H.264 [2]. Coding and carriage parameters are defined in corresponding MISB Engineering Guideline EG 0802, H.264/AVC Coding and Carriage [3].

6.2 Metadata

The approved metadata sets are set forth in EG0104.5 [4] and EG0601.1 [5]. EG0104.5 specifies a mapping for Cursor on Target (CoT) data, which can be mapped for use in EG0601.1. Metadata that is time synchronized to the video is preferred over asynchronous methods. Time stamping of H.264/AVC video is described in RP 0604 [6]. RTP has provisions to carry a local timestamp for each media stream it carries, so although the video and the metadata will maintain independent timestamps the two streams can still be realigned at the decoder.

Metadata consistent with these Engineering Guidelines is encoded using the KLV (Key, Length, Value) construct according to SMPTE 336M-2001 [7].

The approved metadata structure is described in RP 0701, “Common Metadata System: Structure [8]”. Minimum metadata subsets are defined in these documents, although any metadata that is RP 0701 compliant can be used. However, only the approved subset is REQUIRED to be decoded.

7 Streaming Components

7.1 Control

The Real Time Streaming Protocol (RTSP) provides an application level protocol to interactively control the delivery of FMV digital motion imagery delivered via streaming. RTSP is defined in the following specification:

- Real Time Streaming Protocol (RTSP) [9]

RTSP provides a flexible protocol framework for controlling data streams with real-time properties. The following restrictions apply to ensure interoperability between endpoints supporting FMV data streams when this level of control is required or desired:

REQUIRED

- RTSP clients and servers SHALL implement all required features of the minimal RTSP implementation described in Appendix D of RFC 2326.
- RTSP clients and servers SHALL implement the PLAY method.
- RTSP clients and servers SHALL support RTP/AVP transport in the “Transport” header. When the RTP/AVP transport is used for a unicast session, clients SHOULD include the “client_port” parameter in the “Transport” header and servers SHOULD include the “server_port”, “source”, and “ssrc” parameters in the “Transport” header.
- RTSP servers SHALL send the “RTP-Info” header for unicast sessions.
- RTSP servers and clients SHALL support aggregated control of presentations.
- At most one RTSP session SHALL be “active” on a connection between an RTSP client and an RTSP server at any one time. An RTSP session becomes “active” when it is first referenced in a “Session” header. An RTSP session is no longer “active” after a TEARDOWN request has been issued for that session.

RECOMMENDED

- RTSP clients and servers SHOULD implement the DESCRIBE method. If the DESCRIBE method is implemented, it is REQUIRED that SDP be supported as the description format, as specified in Appendix C of RFC 2326.
- RTSP clients SHOULD generate the following RTSP headers when appropriate: “Bandwidth”, “Cache-Control”, “If-Modified-Since”, “User-Agent”. RTSP servers SHOULD correctly interpret these headers when present.
- RTSP servers SHOULD generate the following RTSP headers when appropriate: “Cache-Control”, “Expires”, “Last-Modified”, “Server”. RTSP clients SHOULD correctly interpret these headers when present.

7.1.1 Description and Addressing

Session Description Protocol (SDP) provides a flexible language for describing media streams and relating them temporally. SDP is defined in the following specification:

SDP: Session Description Protocol [10]

The SDP data MUST be formatted according to Appendix C of RTSP (RFC 2326) at all times. Although Appendix C provides compatibility when delivering an SDP that will have media controlled via RTSP it allows for consistent formatting and attribute parsing.

7.1.2 Real Time Control Protocol (RTCP)

RTCP is an optional yet extremely useful companion protocol that provides bi-directional feedback between the sender and the receiver regarding the quality of a RTP session. RTCP allows the sender to provide information to the receiver such as how many bytes and packets have been sent, and it allows the receiver to provide information to the sender such as how many packets were lost, and a measure of the packet arrival jitter.

RTCP is important in synchronizing media streams, such as for lip sync, at the receiver as it carries important time reference information. The update rate of RTCP sender packets is typically 5 s, which means that upon entering a streaming session there may be an initial delay - on average a 2.5 s duration if the default RTCP timing rules are used - when the receiver does not yet have the necessary information to perform inter-stream synchronization. When video and metadata are required to be time synchronized at the receiver RTCP is required.

7.2 Transport

The Real-time Transport Protocol is used to transport (near) real-time digital motion video. RTP provides end-to-end transport for media with real-time characteristics and is widely used in Internet streaming media applications.

The following RFCs provide the core RTP specifications that MUST be implemented. Payload specifications are detailed in following sections.

- RFC 3550 - RTP: A Transport Protocol for Real-Time Applications [11]
- RFC 3551 - RTP Profile for Audio and Video Conferences with Minimal Control [12]

Interleaved RTSP and RTP/AVP over TCP is an OPTIONAL method for transport. This method offers reliable transmission and more easily traverses Network Address Translation (NAT) devices and Firewalls at the expense of real-time time-critical response.

7.2.1 Video

H.264/AVC transport over RTP is defined in [13] with additional guiding parameter elections in [3] and can be used with the following restrictions:

- The interleaved mode (packetization-mode=2) SHALL NOT be used (guarantees lower latency)
- The parameters that are defined for interleaved mode (packetization-mode=2) SHALL NOT be present in the "a=fmtp" line in the SDP
- The parameters **max-mbps**, **max-fs, max-dpb**, **max-br** SHALL NOT be present in the "a = fmtp" line of the SDP
- The format parameters line ("a=fmtp") in the SDP SHALL include the following parameters: **sprop-parameters-sets**, **profile-level-id**

7.2.2 Metadata [TBD]

7.3 Sample RTSP/RTP Session

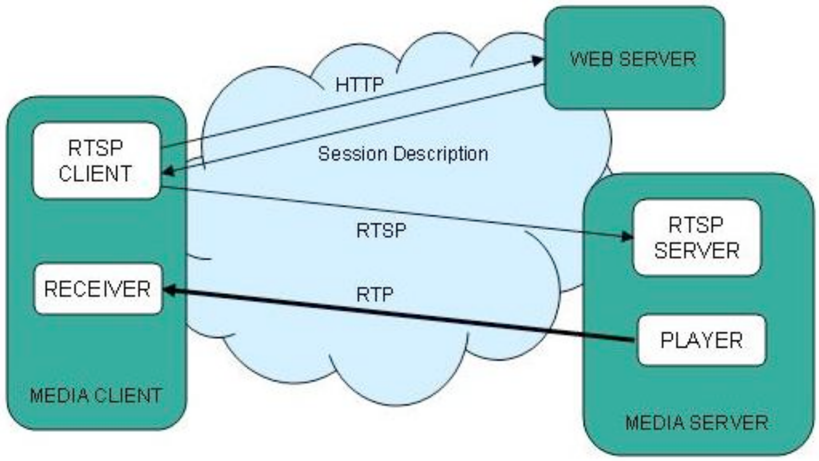

Figure 2. RTSP/RTP Server/Client Communication

The following commands may be used to control a RTP session and are illustrated as state and requests in Figure 3:

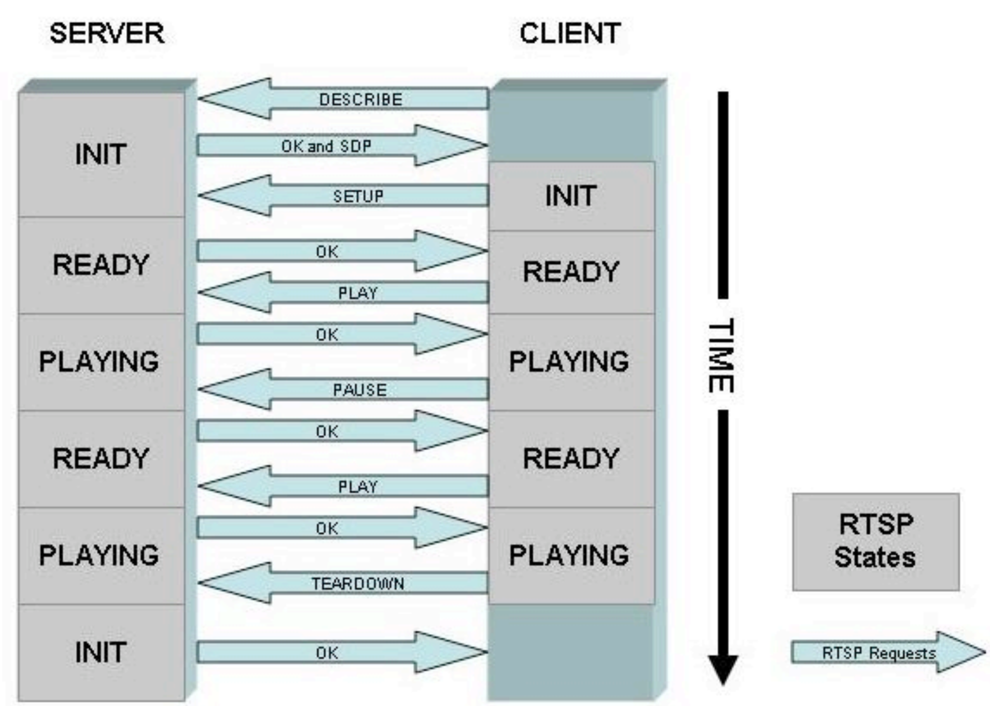

Figure 3. Server/Client RTSP Interaction

DESCRIBE: Used by the client to retrieve a description of a presentation or media object on the server, corresponding to the *Universal Resource Locator* (URL) sent in the request. The response is typically in the form of the *Session Description Protocol* (SDP) and gives details such as the encoding types of the media, media duration, authors, etc. This command allows clients to find out more about a clip prior to streaming and also to check if the client can support the media format.

OPTIONS: Informs the sender what other valid requests it may issue i.e. what requests are supported by the corresponding client or server for the specified content at a given time. Illegal requests by either the client or server can be avoided with this operation.

SETUP: Transport of the requested media is configured using this command. Details such as the transport protocol and the port numbers to use are submitted to the server so the content is delivered in a manner appropriate for the client.

PLAY: Tells the server to start transmitting the requested media content as specified in the associated SETUP command. Unless a SETUP request has been issued and acknowledged, it is illegal to call the PLAY command. The absolute playback time and duration are also specified in this command so operation similar to fast forward and rewind on a VCR can be achieved with this command if the media can support such functionality e.g. live video streams cannot be scanned ahead.

PAUSE: Temporarily interrupts the delivery of the media currently playing in the session. PLAY must have been successfully requested in order to allow pausing of a video stream to a client. Resuming a paused media session is achieved using the PLAY request.

TEARDOWN: Terminates a stream or session. Once this command has been successfully issued the SETUP request must be called again before media content can be streamed again.

Other optional requests defined in the RTSP standard include ANNOUNCE, SET_PARAMETER, GET_PARAMETER, RECORD and REDIRECT. States for each session are maintained by the server to ensure that only valid requests are processed and that an error response is replied to invalid requests. To aid the server in determining if a request is valid a number of possible server states are used:

1. **Init:** the initial state meaning that the server has received no valid SETUP request.

2. **Ready:** the previous SETUP request was successful and acknowledged and the server is waiting to start or finish playing or a valid PAUSE request has been called.
3. **Playing:** a previous PLAY request was successful and content is currently being streamed to the client.

Figure 3 illustrates an example RTSP interaction between a client and server, highlighting the client and server states in response to different RTSP requests. In the session shown, the clients asks for a description of some content contained on the server using the DESCRIBE command and the server delivers information in SDP format with relevant details of the media queried by the client. The client then issues a SETUP request and the server configures the delivery of the stream to suit the clients preferred setup. PLAY is then sent by the client and the server starts transmitting the media data. A PAUSE request then prompts the server to halt the stream temporarily and the server waits in the Ready state for the client issue further instructions. Eventually the client requests PLAY and stream resumes. Finally, the client sends a TEARDOWN request and the server terminates the session and returns to the Init state waiting for other sessions to be established.